EXPRESS MAIL LABEL NO:
EV 302 216 705 US

FABRICATION OF CONDUCTIVE LINES INTERCONNECTING FIRST CONDUCTIVE GATES IN NONVOLATILE MEMORIES HAVING SECOND CONDUCTIVE GATES PROVIDED BY CONDUCTIVE GATE LINES, WHEREIN THE ADJACENT CONDUCTIVE GATE LINES FOR THE ADJACENT COLUMNS ARE SPACED FROM EACH OTHER, AND NON-VOLATILE MEMORY STRUCTURES

BACKGROUND OF THE INVENTION

[0001] The present invention relates to integrated circuits, and more particularly to nonvolatile memories.

[0002] Fig. 1 illustrates an electrically erasable programmable read-only memory array (EEPROM) described in U.S. patent no. 6,420,231 issued July 16, 2002 to Harari et al. and incorporated herein by reference. Fig. 2 is a circuit diagram of the array. Each memory cell 110 has two conductive floating gates 120 positioned side by side in the X direction (row direction) over planar top surface 124T of silicon substrate 124. The floating gates are insulated from the substrate. In the Y direction (column direction), the adjacent floating gates are separated by field oxide regions 130. The floating gates are formed from the first polysilicon layer.

[0003] Steering gates 134 are formed from the second polysilicon layer and are insulated from the floating gates. Each steering gate extends in the Y direction between two columns of memory cells 110 and overlies two adjacent columns of floating gates 120. Bitlines 138 are diffusion regions in substrate 124. Each bitline 138 runs in the Y direction between two adjacent columns of floating gates 120. In each row, a bitline 138 provides two source/drain regions to respective two adjacent memory cells 110.

[0004] Wordlines 144, formed from the third polysilicon layer, overlie the steering gates and extend in the X direction. Wordlines 144 may also be formed from polycide. The wordline layer also provides select gates 144S (Fig. 1) for the memory cells.

[0005] Metal strap lines (not shown) reduce the resistance of the polysilicon elements and diffusion elements of the array.

[0006] The memory operates as follows. Each cell 110 can be represented as having two floating gate transistors 110L, 110R (Fig. 2) separated by a select gate transistor 110S (a transistor with gate 144S). The floating gate of transistor 110L is selected for reading or programming by placing a sufficient voltage on the steering gate 134 above the floating gate of transistor 110R to turn on the transistor 110R regardless of the charge on its floating gate. Likewise, the floating gate of transistor 110R is selected for reading or programming by placing a sufficient voltage on the steering gate 134 above the floating gate of transistor 110L to turn on the transistor 110L regardless of the charge on its floating gate. Each floating gate can be read by providing a voltage difference between the respective bitlines 138 and sensing the state of one of the bitlines. A negative charge can be written to a floating gate by source side hot electron injection. The floating gates can be erased through wordlines 144 or substrate 124. See U.S. patent no. 6,266,278 issued July 24, 2001 to Harari et al. and incorporated herein by reference.

[0007] As noted above, floating gates 120 are made from the first polysilicon layer, steering gates 134 are made from the second polysilicon layer, and wordlines 144 are made from the third polysilicon layer or a polycide layer which also provides the select gates 144S. Alternative fabrication techniques are desirable.

SUMMARY

[0008] This section summarizes some features of the invention. Other features are described in the subsequent sections. The invention is defined by the appended claims which are incorporated into this section by reference.

[0009] In some embodiments of the present invention, select gates 144S are formed from a different layer than wordlines 144. In some embodiments, this fabrication method provides additional control over the spacing between the wordlines 144 and steering gates 134. Increased spacing is desirable to reduce the parasitic capacitance between the wordlines and the steering gates.

[0010] In some embodiments, select gates 144S are formed before floating gates 120 and before steering gates 134 (the steering gates will also be called "control gates" herein). The control gates are provided by conductive control gate lines. Each control gate line provides control gates for one column of the memory cells. The adjacent control gate lines for the adjacent columns are spaced from each other. See Figs. 3A-3G

illustrating one embodiment described in detail below. In some embodiments, wordlines 144 are formed after the floating and control gates. For example, select gates 144S can be formed from the first polysilicon layer, floating gates 120 from the second polysilicon layer, control gates 134 from the third polysilicon layer, and wordlines 144 from a metal layer. The gate dielectric for select gates 144S can be formed by thermal oxidation before the formation of the floating gates. Therefore, the floating gates are not affected (not oxidized) by the select gate oxide growth. Also, in some embodiments, the select gate dielectric has the same thickness as the gate dielectric of high voltage peripheral transistors, so the same gate dielectric layer can be used both for select gates 144S and the high voltage peripheral transistors. See U.S. patent applications no. 10/440,508 filed May 16, 2003 and no. 10/632,154 filed July 30, 2003, both incorporated herein by reference.

[0011] In addition, if the select gates 144S are formed before the floating gates, the gate dielectric for floating gates 120 can be the same layer as the dielectric formed on the select gate sidewalls to insulate the select gates from the floating and control gates. See U.S. patent applications no. 10/440,005 filed May 16, 2003 and no. 10/631,452 filed July 30, 2003, both incorporated herein by reference.

[0012] In some embodiments of the present invention, the floating gates are fabricated in a self-aligned manner using an isotropic etch of the floating gate layer.

[0013] Other features and advantages of the invention are described below. The invention is defined by the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Fig. 1 is a perspective view of a prior art memory array.

[0015] Fig. 2 is a circuit diagram of the array of Fig. 1.

[0016] Fig. 3A is a perspective view of a memory array according to one embodiment of the present invention.

[0017] Figs. 3B, 3C, 3D, 3E show vertical cross sections of the array of Fig. 3A.

[0018] Fig. 3F is a top view of the array of Fig. 3A.

[0019] Fig. 3G is a circuit diagram of the array of Fig. 3A.

[0020] Figs. 4-7, 8A, 8B, 9-12, 13A, 13B, 14A, 14B, 15A show vertical cross sections of memory structures in the process of fabrication according to one embodiment of the present invention.

[0021] Fig. 15B is a top view of a memory structure in the process of fabrication according to one embodiment of the present invention.

[0022] Figs. 15C, 16A, 16B, 16C show vertical cross sections of memory structures in the process of fabrication according to one embodiment of the present invention.

[0023] Fig. 17A is a top view of a memory structure in the process of fabrication according to one embodiment of the present invention.

[0024] Figs. 17B, 17C, 18A, 18B, 18C, 19, 20A, 20B, 21, 22A, 22B, 23A, 23B, 23C, 24A, 24B, 24C, 25A, 25B, 26A, 26B, 27A, 27B, 28A, 28B, 29, 30, 31 show vertical cross sections of memory structures in the process of fabrication according to one embodiment of the present invention.

DESCRIPTION OF SOME EMBODIMENTS

[0025] The embodiments described in this section illustrate but do not limit the invention. The invention is not limited to particular materials, process steps, or dimensions. The invention is defined by the appended claims.

[0026] Figs. 3A-3F are different views of a nonvolatile memory array according to one embodiment of the present invention. Fig. 3A is a perspective view. Fig. 3F is a top view. Figs. 3B, 3C, 3D, 3E show vertical cross sections marked in Fig. 3F as X-B, Y-C, Y-D, and X-E respectively. The cross section X-B runs in the X direction (row direction) through floating gates 120 and the active areas between field dielectric regions 130. The cross section Y-C runs in the Y (column) direction through select gates 144S. The cross section Y-D runs in the Y direction through floating gates 120. The cross section X-E runs in the X direction through oxide regions 130 between select gates 144S. Fig. 3G is a circuit diagram of the array.

[0027] The circuit diagram of Fig. 3G is similar to that of Fig. 2 but each control gate line 134 provides the control gates to one column of the memory cells, and the adjacent

control gate lines are spaced from each other. As in Fig. 1, each memory cell has two conductive floating gates 120 positioned side by side in the X direction over the planar top surface of active areas 312 of silicon substrate 124. The floating gates are insulated from the substrate by dielectric 314. In the Y direction, the adjacent floating gates are separated by field oxide regions (substrate isolation regions) 130. In this embodiment, the substrate isolation is STI type (shallow trench isolation). Oxide 130 is formed in trenches 130TR. Each trench 130TR runs through the whole array, but oxide 130 is etched out of the trenches at the location of bit lines 138 (note Fig. 3E showing a cross section along a trench 130TR). Oxide 130 protrudes upward above the top surface 124T of substrate 124.

[0028] The invention is not limited to STI. For example, the oxide 130 structure of Fig. 1 can also be used.

[0029] In Figs. 3A-3F, control gates 134 overlie floating gates 120, and are insulated from the floating gates and from select gates 144S by ONO 324. ONO 324 (oxide/nitride/oxide) is a sandwich of silicon dioxide, silicon nitride, silicon dioxide. Each control gate line 134 extends in the Y direction and overlies one column of floating gates 120. The adjacent control gate lines are spaced from each other. Bitlines 138 are diffusion regions in substrate 124. Each bitline 138 runs in the Y direction between two adjacent columns of floating gates 120, traversing the trenches 130TR. In each row except the first and the last rows of the array, a bitline 138 provides two source/drain regions to two adjacent memory cells.

[0030] In each memory cell, a channel region in the active area of substrate 124 extends between the two adjacent bitlines. The channel region includes two channel areas underlying two respective floating gates 120 and a channel area underlying the select gate 144S of the cell.

[0031] In this embodiment, control gates 134 are doped polysilicon, but polycides and other conductive materials can also be used.

[0032] Conductive select gates 144S are insulated from substrate 124 by dielectric 330. In this embodiment, select gates 144S are doped polysilicon, but they may be made of polycides or other conductive materials. Each select gate 144S may overlap the adjacent STI trenches 130TR (Figs. 3C, 3F), but each gate 144S does not extend to the adjacent memory cells. The wordlines 144 (e.g. metal) are formed from a separate layer.

Before that layer is deposited, dielectric layers 302, 304, 310 are formed over the structure. The combined thickness of layers 302, 310 and the combined thickness of layers 304, 310 can be large to reduce the capacitance between the wordlines 144 and the underlying memory features (including the floating and control gates). In some embodiments, dielectric 302 is 1000-1500Å thick, dielectric 304 is 200-400Å thick, and dielectric 310 is 2000-5000Å thick, and larger thickness values can be used. Contact openings 340 are etched in ILD 310 to select gates 144S. The openings are filled with metal 350. Metal 350 may be part of wordline layer 144, or may be a separate layer (e.g. tungsten plugs).

[0033] The term "control gate" may apply to a control gate of a single memory cell or to a line 134 running through the array.

[0034] Exemplary operating voltages are shown in Table 1 below. The voltages are in volts. The reading and programming operations are performed as in the memory of Fig. 1. The programming is by channel hot electron injection (CHIE). The erase is through substrate 124 (Fowler-Nordheim tunneling through dielectric 314). The sign "~"indicates a voltage range. For example "1~2" means 1 V to 2 V.

[0035] TABLE 1

			Read	**Program (CHEI)**	**Erase**
WL 144 (Row)	Selected		2.5	1.5	1~2
	Unselected		0	0	0
CG 134	Selected Column	Selected bit (L or R)	1.5~2	9~10	-9 ~ -10
		Unselected bit (R or L)	6~7	6~7	
	Unselected Column		0	0	
BL 138	Selected	Selected bit (L or R)	1~1.5	4.5~5	Floating

			Read	Program (CHEI)	Erase
	Column	Unselected bit (R or L)	0	0	
	Unselected Column		0	0	
Substrate 124			0	0	7~8

[0036] The programming can also be done by Fowler-Nordheim tunneling.

[0037] In one embodiment, the memory is fabricated as follows. The memory array is fabricated in a doped region of type P- formed in monocrystalline silicon substrate 124. This region can be isolated by P-N junctions (not shown). See e.g. U.S. patent no. 6,355,524 issued March 12, 2002 to H.T. Tuan et al. and incorporated herein by reference.

[0038] STI trenches 130TR can be formed, for example, by a process described in U.S. patent application no. 10/678,317 US filed October 3, 2003 by Yi Ding and incorporated herein by reference. More particularly, as shown in Fig. 4, silicon dioxide 410 (pad oxide) is formed on substrate 124 by thermal oxidation or some other technique to an exemplary thickness of 150 Å. The dimensions, and the voltages in Table 1, are given for an exemplary process using a 0.18 μm line width technology (the line width is the minimal dimension that can be reliably printed photolithographically). The fabrication method is believed to be scalable to smaller line widths (e.g. 90 nm or even smaller), and the invention is not limited to a particular line width. Fig. 4 shows a Y cross section, i.e. a cross section in the Y direction. All the Y cross sections are identical at this stage.

[0039] Silicon nitride 420 is deposited on oxide 410. In one embodiment, the thickness of nitride 420 is in the range of 2000 Å to 2200 Å. Nitride 420 is patterned photolithographically, using a photoresist mask (not shown), to define trenches 130TR and active areas 312. Oxide 410 and substrate 124 are etched through the openings in nitride 420. Trenches 130TR ("STI trenches") are formed as a result. An exemplary depth of trenches 130TR is 0.2 μm, measured from the top surface of the substrate 124. Other depths are possible.

[0040] Nitride 420 is subjected to a wet etch to recess the vertical edges of the nitride layer away from trenches 130TR. See Fig. 5 (Y cross section). This step reduces the aspect ratio of the holes that will be filled with dielectric 130. As will be seen below, the etch of nitride 420 will also reduce the capacitance between the floating gates and the substrate 124, thus increasing the gate coupling ratio.

[0041] A thin layer 130.1 of silicon dioxide is thermally grown on the exposed silicon surfaces to round the edges of trenches 130TR and passivate the trench surfaces. Silicon dioxide 130.2 (Fig. 6, cross section Y) is deposited by a high density plasma process (HDP) or by non-plasma low pressure chemical vapor deposition (LPCVD). Oxide 130.2 fills the trenches and initially covers the nitride 420. Oxide 130.2 is polished by CMP (chemical mechanical polishing). The CMP stops on nitride 420. A planar top surface is provided.

[0042] In some of the figures, the layers 130.1, 130.2 are shown as a single layer 130. This dielectric 130 will be referred to as STI dielectric or field dielectric.

[0043] The array area is masked with photoresist (not shown), and oxide 130 is etched to lower the oxide level in the periphery. See Fig. 7 showing the Y cross section of the array and a peripheral cross section. The lower oxide level will increase the peripheral surface planarity, thus facilitating the subsequent fabrication steps. In some embodiments, the top level of oxide 130 is lowered to a level of 200-500 Å above the substrate 124 in the periphery.

[0044] The photoresist is removed. Nitride 420 is removed selectively to oxide 130. This can be done by a wet etch (e.g. with phosphoric acid). See Fig. 8A (cross section Y) and Fig. 8B (periphery). Then pad oxide 410 (Fig. 7) is etched away to expose the substrate 124 in the active areas. The oxide etch may remove a small amount of oxide 130.

[0045] Silicon dioxide 330 is thermally grown on the exposed areas of substrate 124 to provide gate dielectric for the select gates of the memory array and for the peripheral transistors. The peripheral transistors are used to form sense amplifiers, address decoders, memory I/O buffers, drivers for various memory elements (e.g. bitlines, wordlines, control gates), and possibly other circuitry needed to access the memory array. See e.g. the aforementioned U.S. patents 6,420,234 and 6,266,278. An exemplary

thickness of oxide 330 in the array area is 120 Å. Generally, the oxide thickness depends on the maximum voltage that the oxide 330 is designed to sustain during the memory operation.

[0046] In the example shown in Fig. 8B, the peripheral area includes a high voltage transistor area 812H and a low voltage transistor area 812L. Oxide 330 is first grown thermally to a thickness of 60 Å over the entire wafer. This oxide is removed from the low voltage area 812L by a masked etch. The wafer is re-oxidized to re-grow silicon dioxide 330 in area 812L to a thickness of 60 Å. The oxide thickness in the memory array area and in high voltage area 812H increases from 60 Å to 120 Å during this step.

[0047] Thus, oxide 330 in the array area and oxide 330 in the high voltage peripheral area 812H is formed simultaneously in these two oxidation steps. The oxide 330 in area 812L and the oxide 330 in the array area and area 812H are not formed simultaneously because the oxide 330 in area 812L is formed in the second oxidation step. See U.S. patent application no. 10/440,508 filed May 16, 2003 by Yi Ding and incorporated herein by reference.

[0048] Intrinsic polysilicon 144.1 (Fig. 9, cross section Y and peripheral cross section) is deposited over the wafer. Layer 144.1 will provide portions of select gates 144S and peripheral transistor gates. An exemplary deposition process is LPCVD, and an exemplary thickness is 1000-1400Å. Polysilicon 144.1 fills the recesses between the STI dielectric regions 130 and covers the whole wafer.

[0049] Silicon dioxide 902 is deposited on polysilicon 144.1 to an exemplary thickness of 1200-1500Å by CVD (TEOS) or some other process.

[0050] A photoresist mask 904 is formed over the periphery. Oxide 902 is etched away in the array area, and polysilicon 144.1 is doped N+ by ion implantation. See Fig. 10 (cross section Y and peripheral cross section). Alternatively, the doping can be performed before the etch of oxide 902, by ion implantation through the oxide. Resist 904 blocks the dopant from the periphery. Polysilicon 144.1 remains undoped in the periphery, and will be doped later together with the peripheral source/drain regions to create surface channel peripheral transistors. These details are exemplary and not limiting. Non-surface-channel transistors can also be used in the periphery, and further the surface channels transistors can be created by other techniques, known or to be

invented.

[0051] Polysilicon 144.1 is subjected to a timed dry anisotropic etch to lower its top surface below the top surface of dielectric 130 in the array area. See Fig. 11 (cross section Y and peripheral cross section). Resist 904 protects the periphery during this etch.

[0052] Oxide 130 is etched selectively to polysilicon 144.1 (Fig. 12, cross section Y). The etch includes a lateral component that causes the sidewalls of oxide 130 to be laterally recessed in the direction away from the adjacent polysilicon features 144.1 and active areas 312. This can be an isotropic wet etch. In one embodiment, the isotropic etch laterally recesses the sidewall by an amount Ls in the range of 0.04 μm to 0.05 μm. The etch also lowers the top surface of oxide 130 by the same amount. In addition, the etch attacks the oxide portions near the polysilicon 144.1 to form pockets 1210 in which the top surface of oxide 130 is below the top surface of polysilicon 144.1. Between the pockets 1210, the top surface of oxide 130 is shown to be above the top surface of polysilicon 144.1, but this is not necessary. The top surface of oxide 130 between the pockets 1210 may be even with, or below, the top surface of polysilicon 410.1. See U.S. patent application no. 10/678,317 filed October 3, 2003 by Yi Ding and incorporated herein by reference.

[0053] The periphery is protected by resist 904, and remains as in Fig. 11.

[0054] As seen in Fig. 3D, the recessed sidewalls of oxide 130 will allow the top surface of floating gates 120 to extend over the oxide 130, possibly beyond the active areas 312, advantageously increasing the capacitive coupling between the floating and control gates.

[0055] Resist 904 is removed. Polysilicon layer 144.2 (Fig. 13A, cross section Y, and Fig. 13B, periphery) is deposited on the structure, and is doped N+ during or after the deposition. Layer 144.2 will provide portions of the select gates 144S. An exemplary deposition process is conformal low pressure chemical vapor deposition (LPCVD). The thickness of layer 144.2 (at least 2000 Å in some embodiments) is chosen to provide a planar top surface. The planar top surface is not necessary however.

[0056] Polysilicon layers 144.1, 144.2 are shown as a single layer 144S in the array area in some of the drawings.

[0057] A hard mask for an etch of polysilicon layers 144.1, 144.2 is formed in two steps. First, silicon nitride layer 1310 is deposited and patterned photolithographically to form the same pattern in the array area as was used to define active areas 312 (Fig. 4). Nitride 1310 covers the entire periphery. Next, a conformal silicon nitride layer 1410 (Fig. 14A, cross section Y) is deposited and etched anisotropically without a mask to form spacers on the sidewalls of nitride 1310. Nitride 1410 is etched off the periphery during this step, so the periphery remains as in Fig. 13B. Nitride spacers 1410 will ensure that the select gates 144S will overlap the top planar surface of STI oxide 130 even if nitride 1310 is not perfectly aligned with active areas 312.

[0058] Alternatively, nitride 1310 can be removed from the periphery when the nitride 1310 is patterned in the array area. Nitride 1410 will also be removed from the periphery during the etch that forms the nitride spacers in the array area. The resulting peripheral cross section for this case is shown in Fig. 14B.

[0059] An anisotropic etch of polysilicon 144S (Fig. 15A, cross section Y, and Fig. 15B, top view of the array) stops on STI oxide 130. Some of oxide 130 may be removed by an overetch. Polysilicon 144S forms a number of strips (Fig. 15B); each strip runs through the memory array in the X direction over an active area 312. (Nitride layers 1310, 1410 are not shown in Fig. 15B.)

[0060] If the periphery was as in Fig. 13B, it will not be affected by the polysilicon etch. If the periphery was as in Fig. 14B, polysilicon 144.2 will be removed from the periphery, as shown in Fig. 15C.

[0061] Advantageously, the array structure is essentially self-aligned at this stage. The positions of trenches 130TR, oxide 130 and polysilicon 144S does not depend on photolithographic alignment except for the alignment of the mask used to pattern the nitride 1310 with respect to the mask used to pattern the trenches 130TR (Fig. 4). A misalignment between these two masking steps may cause a shift of polysilicon strips 144S relative to oxide 130 in the Y direction, but the shift is not believed to affect the memory characteristics if nitride spacers 1410 are sufficiently wide to ensure the overlap of the polysilicon strips onto the top surface of oxide 130 (as in Fig. 15A).

[0062] Nitride layers 1310, 1410 are removed from the wafer (by a wet H_3PO_4 etch for example). Silicon nitride 1610 (Fig. 16A, cross section Y) is deposited over the

structure to provide a planar top surface (the planar top surface is not necessary however). The peripheral cross section is shown in Figs. 16B, 16C for the cases of Figs. 13B, 15C respectively. An exemplary thickness of nitride 1610 over the top surface of polysilicon 144S is in the range of 500-1500Å. A SION or silicon dioxide layer 1620 is formed on nitride 1610 to provide additional protection for a subsequent polysilicon etch that will define the select gates 144S. Layer 1620 is not needed if nitride 1610 is sufficiently thick (1500Å for example).

[0063] Layers 1610, 1620 are patterned photolithographically using a photoresist layer (not shown) to form a number of strips running through the array in the Y direction over the future positions of select gates 144S. See Fig. 17A (top view of the array). The photolithographic step is not alignment-sensitive in the array area because any misalignment will simply shift the strips 1610, 1620 in the X direction and will not affect the array geometry. Layers 1610, 1620 are removed from the periphery during this step, as shown in Fig. 17B (for the case of Fig. 16B) and Fig. 17C (for the case of Fig. 16C).

[0064] The resist used to pattern the layers 1610, 1620 can be removed at this stage, or can be left in the structure for the etch of polysilicon 144S. Polysilicon 144S is etched anisotropically to form the select gates 144S. The etch is selective to SION 1620 and nitride 1610, or at least to the resist if the resist is present. See Fig. 18A showing the cross section X-B (marked in Fig. 3F), and Fig. 18B showing the cross section X-E. A dashed line in Fig. 18A shows the position of the top surface 130T of STI oxide 130 between select gates 144S (see also Fig. 16A). In the case of Fig. 17B, polysilicon 144.2 is removed from the periphery during this etch, so the periphery becomes as in Fig. 17C. See also Fig. 18C, showing the periphery both for the case of Fig. 17B and the case of Fig. 17C after the polysilicon etch.

[0065] If the resist was present during the polysilicon etch, the resist is removed. The wafer is oxidized (in a furnace or by rapid thermal oxidation (RTO)) to grow silicon dioxide 314 (Fig. 18A) on substrate 124. During this step, silicon dioxide 1820 grows on the sidewalls of polysilicon gates 144S in the array area, and a thin oxide layer 1830 grows on the vertical sidewalls of nitride 1610. Oxide 1830 may also form on the top surface of nitride 1610 if SION 1620 is omitted. Oxide 314 will serve as the tunnel oxide (the gate oxide for the floating gate transistors). Oxide 1820 will provide sidewall insulation for the select gates. The oxide thickness depends on the dopants and dopant

concentrations, and is chosen based on a desired memory cell performance. In one embodiment, oxide 314 is 60-100 Å thick, and oxide 1820 is 250-450 Å thick (due to the heavier doping of select gates 144S).

[0066] The peripheral area is covered by oxide 902 (Fig. 18C), and is not affected by the oxidation.

[0067] Floating gate polysilicon 120 (Fig. 19, cross section X-B) is deposited over the structure, by LPCVD for example, and is doped during or after the deposition. Polysilicon 120 is sufficiently thick to ensure that its top surface is at least as high throughout the wafer as the surface 130T (Fig. 16A) of STI oxide 130. In the embodiment of Fig. 19, the polysilicon surface is planar except for protrusions over the select gates 144S. The surface 120T of polysilicon 120 between the STI regions 130 is planar due to a conformal polysilicon deposition if the maximum distance D1 (Figs. 3A, 3D) between adjacent STI oxide regions 130 above substrate 124 is not larger than twice the thickness Th1 (Fig. 19) of polysilicon 120, i.e. $D1 \leq 2*Th1$. The polysilicon protrusions over select gates 144S (Fig. 19) are obtained if twice the thickness Th1 of polysilicon 120 is smaller than the distance D2 between the adjacent structures consisting of select gates 144S and oxide 1820 (see also Fig. 18A). In summary,

$$D1 \leq 2*Th1 < D2.$$

[0068] In some 0.18μm minimum line width embodiments, D1 is 0.16-0.2μ, D2 is 0.58μm, and Th1 is at least 0.08μm but less than 0.29μm.

[0069] In other embodiments, the polysilicon surface is planar over the whole wafer due to a conformal deposition to a thickness larger than half the distance D2 and larger than half the distance D1. See U.S. patent application no. 10/440,508 filed May 16, 2003 by Yi Ding and incorporated herein by reference.

[0070] In other embodiments, the polysilicon 120 is not planar even in the areas between select gates 144S. For example, polysilicon 120 can have protrusions over STI oxide regions 130.

[0071] Polysilicon 120 is etched without a mask. In the embodiment of Fig. 19, this can be an isotropic dry etch. The etch end point is the exposure of oxide surface 130T. See Figs. 20A (cross section X-B) and 20B (cross section Y-D). The etch removes

polysilicon 120 over the select gates 144S, and also removes SION 1620. The remaining polysilicon 144S fills the areas between protruding oxide features 130 and select gates 144S. Oxide layers 1820, 1830 become exposed. Oxide 1830 protects the nitride 1610 during the etch, and may be partially or completely removed by this etch. The polysilicon etch removes polysilicon 120 from the periphery, so the periphery becomes as in Fig. 18C.

[0072] Anisotropic etches can also be used, especially if polysilicon 120 is initially planar throughout the wafer. If the polysilicon is not planar, it can be planarized before the anisotropic etch.

[0073] An optional etch of oxide 130 lowers the top surface of oxide 130 to a level below the top surface of polysilicon 120 (Fig. 21, cross section Y-D) to increase the capacitive coupling between floating gates 120 and control gates 134 (Fig. 3D). See the aforementioned U.S. patent no. 6,355,524. The oxide etch can be an anisotropic dry etch. Some of oxide 902 can be removed in the periphery by this etch (see Fig. 18C).

[0074] As noted above in connection with Fig. 12, the recessed sidewalls of oxide 130 allow the floating gates to be wider at the top (by the amount Ls) and to possibly extend beyond the active areas 312. See also Figs. 3A, 3B, 3D. The lateral extensions of the floating gates increase the capacitive coupling between the floating gates and the control gates 134 and increase the gate coupling ratio. Also, due to the lateral extensions, the spacing between the floating gates is reduced, possibly below the minimum line width, thus allowing a more efficient use of the wafer area. In one embodiment, the spacing Sf (Figs. 3D, 21) between the adjacent floating gates is only 0.05-0.06 μm for the dimensions given above in connection with Fig. 12.

[0075] The wafer is cleaned for subsequent processing. The cleaning step removes any oxide 1830 (Figs. 18A, 18B, 20A) that may have remained in the structure after the isotropic etch of polysilicon 120.

[0076] Insulating layer 324 (Fig. 22A, cross section X-B, Fig. 22B, cross section X-E, and Fig. 22C, periphery), e.g. ONO, is formed over the structure. A conductive layer 134, e.g. doped polysilicon, is deposited over ONO 324. In the embodiment shown, layer 134 is deposited conformally to a thickness of 1000-1800Å, although neither the conformal deposition nor the thickness range are necessary. Polysilicon 134 forms

protrusions 134P over the select gates 144S and nitride features 1610. Each protrusion 134P is a continuous ridge running through the array over one column of select gates 144S and nitride features 1610 in the Y direction. Protrusions 134P will be exploited to form the control and floating gates in a self-aligned manner, as described below. The width W1 of polysilicon 134 over the sidewalls of gates 144S will define the width of the control and floating gates. In case of a conformal deposition, W1 is equal to the thickness of layer 134.

[0077] As shown in Figs. 22A and 22B, cavities 134C form in layer 134 between protrusions 134P. These cavities are filled with some material 2210. In one embodiment, material 2210 is silicon dioxide deposited on polysilicon 134 and planarized by CMP or some other process. The CMP stops when the polysilicon 134 becomes exposed, and then a short wet etch (wet dip) of oxide 2210 is performed to ensure that the oxide 2210 will have vertical sidewalls (the oxide could initially be rounded due to the rounded profile of polysilicon 134). Alternatively, the CMP can stop on ONO 324 and/or nitride 1610.

[0078] In the embodiment of Fig. 23C, the CMP removes the oxide 2210 from the peripheral area, although this is not necessary.

[0079] Polysilicon 134 is etched without a mask selectively to oxide 2210. See Fig. 23A (cross section X-B), Fig. 23B (cross section X-E), and Fig. 23C (periphery). This can be an anisotropic etch which attacks the polysilicon protrusions 134P and creates cavities 2310 in the top surface of the structure. Each cavity 2310 is a continuous cavity running through the array in the Y direction. Polysilicon 134 is recessed relative to oxide 2210 in these cavities. This etch exposes ONO 324 on top of select gates 144S and nitride features 1610. In the embodiment of Figs. 23A, 23B, this etch continues for some time after the exposure of ONO 324 to recess the polysilicon 134 below the top surface of ONO 324. This is not necessary however. The final height Hcg of polysilicon 134 in cavities 2310 (near the select gates 144S) will define the thickness of the control gates. In some embodiments, Hcg is about 500-1000Å, and is less than the thickness of polysilicon 134 under oxide 2210.

[0080] The polysilicon etch removes polysilicon 134 from the periphery (Fig. 23C).

[0081] Dielectric 302 is deposited into cavities 2310 to protect the polysilicon 134 (the future control gates) near the select gates 144S and nitride features 1610. In one

embodiment, this material is silicon nitride. As shown in Figs. 23A, 23B, 23C, nitride 302 can be deposited over the structure conformally to a thickness sufficient to obtain a planar top surface over the array and the periphery. A planar top surface is not necessary however.

[0082] Nitride 302 is etched to expose oxide 2210. See Figs. 24A (cross section X-B), 24B (cross section X-E), and 24C (periphery). This can be a dry or wet etch or a CMP. The etch can be terminated before the exposure of ONO 324. The etch can reduce the thickness of nitride 302 in the periphery or remove the nitride. In the example of Fig. 24C, the nitride is removed.

[0083] The periphery is covered by photoresist (not shown). Oxide 2210 is removed (by a wet etch for example). See Figs. 25A (cross section X-B), 25B (cross section X-E).

[0084] Polysilicon 134 is etched anisotropically in the array area with nitride 302 as a mask. The polysilicon etch is selective to silicon dioxide, so the etch stops on ONO 324. The resulting structure is shown in Figs. 26A (cross section X-B), 26B (cross section X-E). This etch patterns the control gate lines.

[0085] The periphery continues to be covered by the resist (not shown) deposited before the etch of oxide 2210, and the periphery thus remains as in Fig. 24C.

[0086] ONO 324 and polysilicon 120 are etched with nitride 302 as a mask to pattern the floating gates. Layers 324, 120 are completely removed from the areas not covered by nitride 302. See Figs. 27A (cross section X-B), 27B (cross section X-E). The periphery continues to be covered by the resist, and is unchanged. In the array area, nitride 302 and STI oxide 130 can be partially removed during the etch of ONO 324. In some embodiments, the ONO etch may also remove the ONO over the top of nitride 1610 and attack the nitride 1610.

[0087] Then bitlines 138 are created by a "self-aligned source" technique. More particularly, as the peripheral area continues to be masked with the resist (not shown), a silicon dioxide etch selective to silicon nitride removes the exposed oxide 130 from trenches 130TR. See Fig. 28A (cross section X-E) and Fig. 3E. The exposed portions of oxide 314 are also removed by this etch (Fig. 28B, cross section X-B). Substrate 124 becomes exposed in the bitline areas 138 (Fig. 3F). Dopant is implanted into these areas to form the bitlines. In an exemplary embodiment, a combination of a shallow arsenic

implant and a deep phosphorus implant are used for the bitlines.

[0088] The resist is removed from over the periphery. The periphery remains as in Fig. 24C.

[0089] The structure is oxidized to grow a silicon dioxide layer 2910 (Fig. 29, cross section X-B, and Fig. 3E) on the sidewalls of floating gates 120 and control gates 134 and on the exposed portions of substrate 124. This step repairs the damage to oxide 314 caused by the etch of polysilicon 120 that formed the floating gates (Fig. 27A). Oxide 2910 is a thin layer whose thickness depends on the doping of the underlying silicon surfaces. The periphery remains as in Fig. 24C.

[0090] The array is covered with photoresist (not shown). ONO 324 and oxide 902 (Fig. 24C) are removed in the periphery (possibly with a dry etch). The resist is removed, and another photoresist layer (not shown) is formed to define the peripheral transistor gates. Polysilicon 144.1 is patterned in the periphery to form these transistor gates (Fig. 30). The resist is removed. Then LDD (lightly doped drain) extensions for peripheral source/drain regions 3008 are formed by ion implantation for the PMOS and NMOS transistors using appropriate photoresist masks (not shown). A thin pad layer 3010 of silicon dioxide is grown on the top and sidewall surfaces of peripheral gates 144.1 using known techniques.

[0091] Silicon nitride spacers 304 (Fig. 31, periphery, and Figs. 3B and 3E) are formed by a conformal deposition and an anisotropic etch of silicon nitride. Then N+ and P+ implants are performed to finish the doping of source/drain regions 3008 for the peripheral transistors. The peripheral PMOS gates 144.1 are doped P+, and the peripheral NMOS gates are doped N+, during the source/drain doping. Surface channel transistors are formed in the periphery as a result. Appropriate masking steps are used to block these implants from the wafer areas in which the doping is undesirable. Then a wet silicon dioxide etch is performed to remove the exposed portions of oxide 3010 (Fig. 30) over the peripheral silicon source/drain regions 3008. A metal (e.g. cobalt) is deposited over the wafer, and the wafer is heated to form conductive metal silicide layers 144L, 3008L on peripheral gates 144.1 and source/drain regions 3008 respectively. The unreacted metal is removed.

[0092] The remaining processing is conventional. In one example, interlevel

dielectric 310 is formed over the structure, to an exemplary thickness of 2000-5000Å. Contact openings are etched in dielectric 310 and nitride 304 to select gates 144S, control gates 134, peripheral transistor gates 144.1 (i.e. to silicide 144L), and source/drain regions 3008 (i.e. to silicide 3008L). The contact openings are filled with metal 350, then metal 144 is deposited and patterned to form the wordlines and perhaps other features.

[0093] The invention is not limited to the embodiments described above. The invention is not limited to the dimensions, materials or voltages shown, or to STI, silicidation, or other processes. Other techniques can be used to pattern the floating and control gates. See for example U.S. patent application no. 10/393,202 filed by Yi Ding on March 19, 2003, and U.S. patent application no. 10/631,941 filed by Yi Ding on July 30, 2003, both incorporated herein by reference. In Figs. 3A-3F, wordlines 144 are perpendicular to control gate lines 134 and bitlines 138, but this is not necessary. The invention is applicable to non-silicon semiconductor memories. The invention is not limited to a memory erased through the substrate, or to any reading, erase or programming methods. The invention covers both flash and non-flash memories. Other embodiments and variations are within the scope of the invention, as defined by the appended claims.